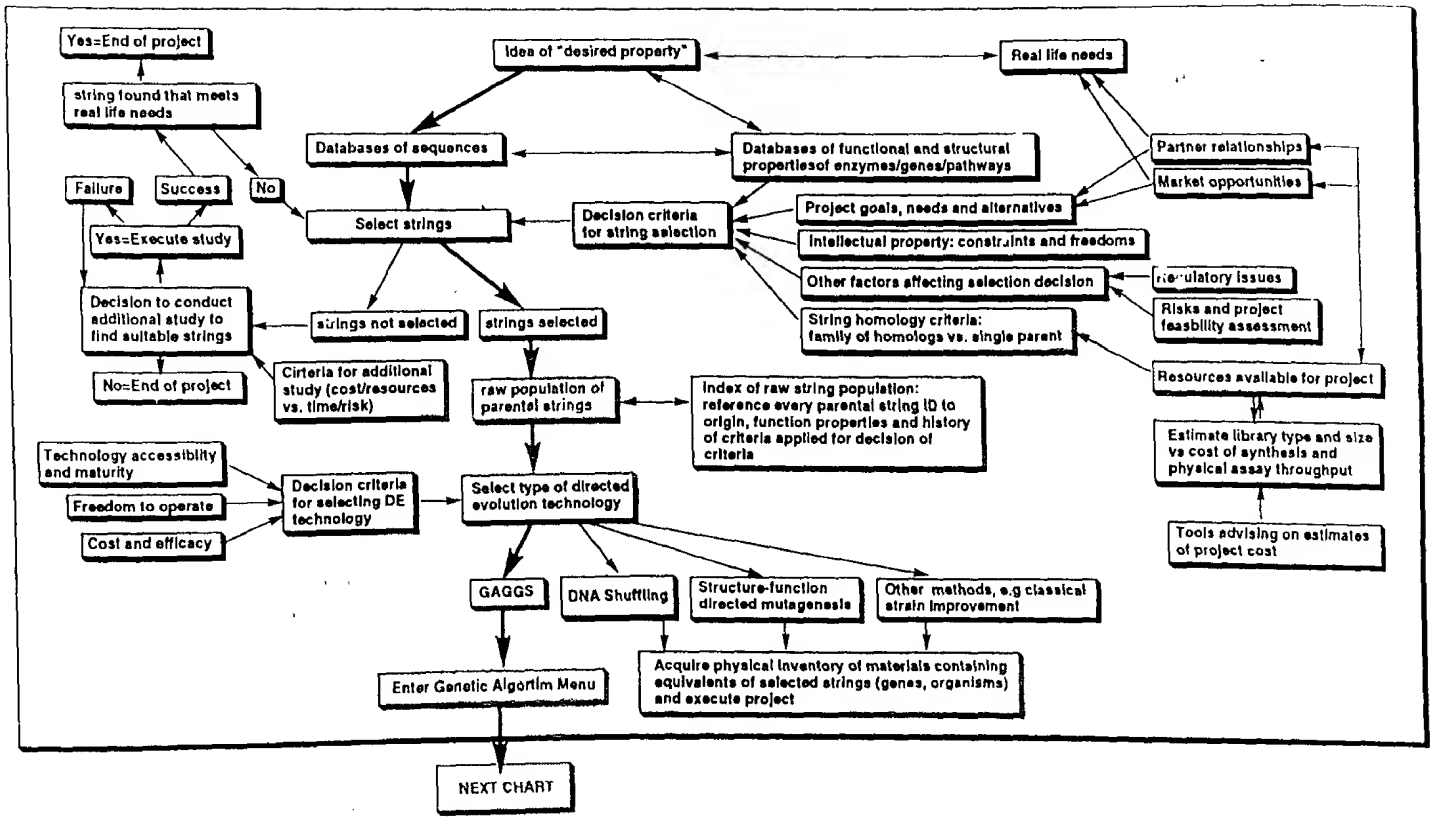
Figure 1


Yes=End of project
string found that meets real life needs
Failure
Success
No
Yes=Execute study
Decision to conduct additional study to find suitable strings
No=End of project
Criteria for additional study (cost/resources vs. time/risk)
Idea of "desired property"
Real life needs
Databases of sequences
Databases of functional and structural propertiesof enzymes/genes/pathways
Select strings
Decision criteria for string selection
strings not selected
strings selected
raw population of parental strings
Index of raw string population: reference every parental string ID to origin, function properties and history of criteria applied for decision of criteria
Partner relationships
Market opportunities
Project goals, needs and alternatives
Intellectual property: constraints and freedoms
Other factors affecting selection decision
String homology criteria: family of homologs vs. single parent
Regulatory issues
Risks and project feasibility assessment
Resources available for project
Estimate library type and size vs cost of synthesis and physical assay throughput
Tools advising on estimates of project cost
Technology accessibility and maturity
Freedom to operate
Cost and efficacy
Decision criteria for selecting DE technology
Select type of directed evolution technology
GAGGS
DNA Shuffling
Structure-function directed mutagenesis
Other methods, e.g classical strain improvement
Enter Genetic Algorithm Menu
Acquire physical inventory of materials containing equivalents of selected strings (genes, organisms) and execute project
NEXT CHART

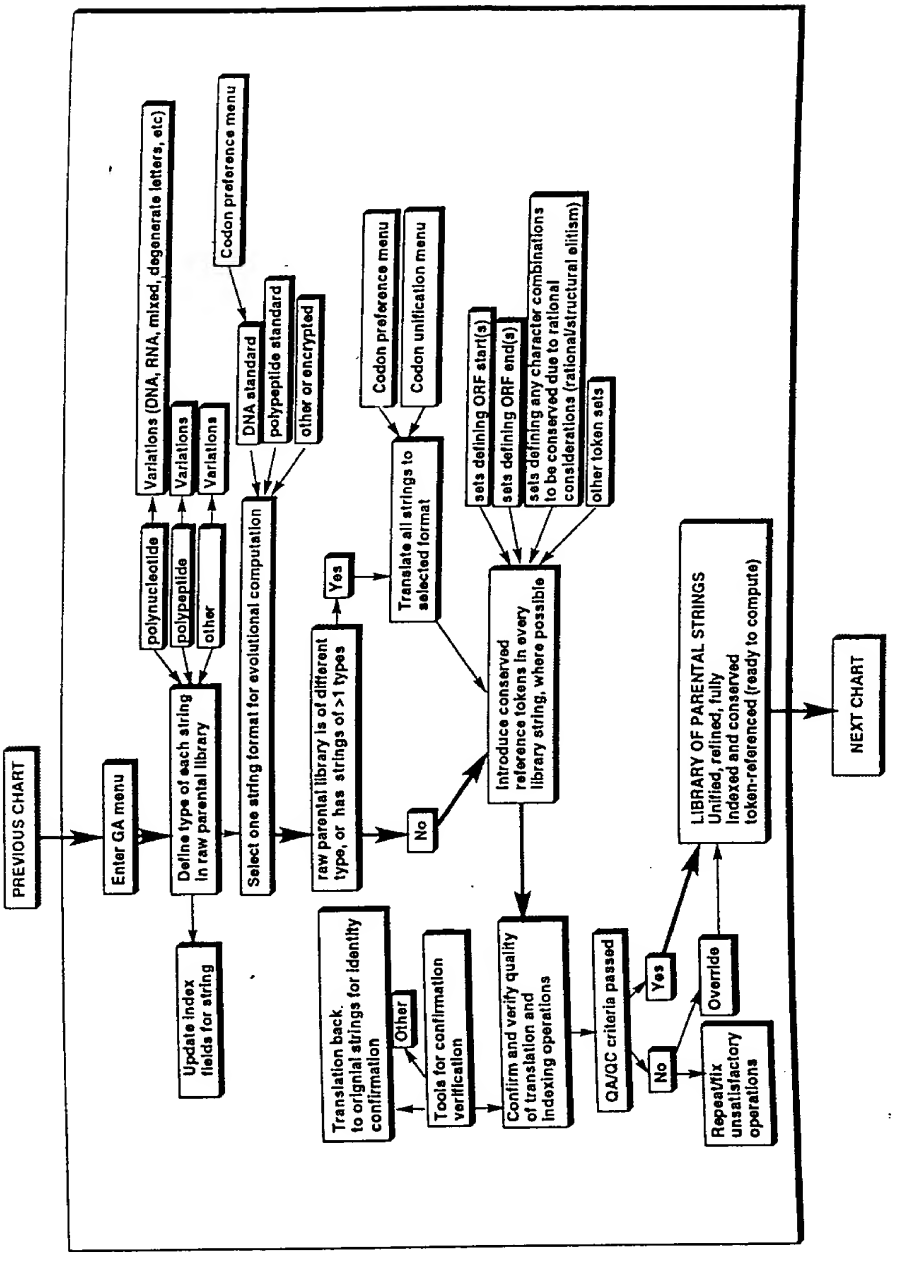

PREVIOUS CHART
Enter GA menu
Define type of each string in raw parental library
Update index fields for string
polynucleotide
polypeptide
other
Variations (DNA, RNA, mixed, degenerate letters, etc)
Variations
Variations
Select one string format for evolutional computation
Codon preference menu
DNA standard
polypeptide standard
other or encrypted
raw parental library is of different type, or has strings of >1 types
Yes
No
Translate all strings to selected format
Codon preference menu
Codon unification menu
Introduce conserved reference tokens in every library string, where possible
sets defining ORF start(s)
sets defining ORF end(s)
sets defining any character combinations to be conserved due to rational considerations (rational/structural elitism)
other token sets
Translation back to orignial strings for identity confirmation
Other
Tools for confirmation verification
Confirm and verify quality of translation and indexing operations
QA/QC criteria passed
Yes
No
Override
Repeat/fix unsatisfactory operations
LIBRARY OF PARENTAL STRINGS
Unified, refined, fully indexed and conserved token-referenced (ready to compute)
NEXT CHART

Figure 2

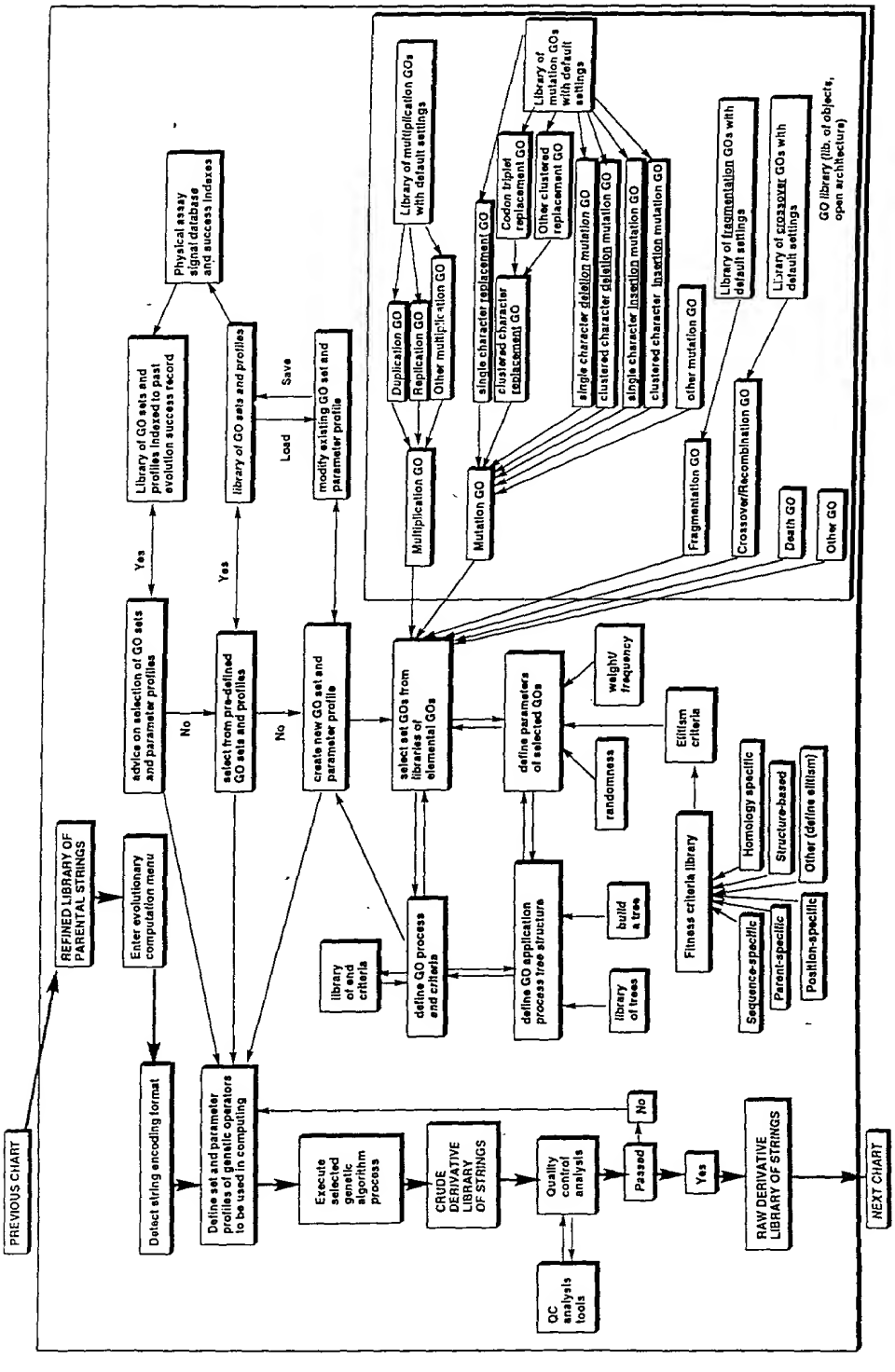

PREVIOUS CHART
REFINED LIBRARY OF PARENTAL STRINGS
Enter evolutionary computation menu
Detect string encoding format
Define set and parameter profiles of genetic operators to be used in computing
Execute selected genetic algorithm process
CRUDE DERIVATIVE LIBRARY OF STRINGS
Quality control analysis
QC analysis tools
Passed
No
Yes
RAW DERIVATIVE LIBRARY OF STRINGS
NEXT CHART
advice on selection of GO sets and parameter profiles
Yes
No
Library of GO sets and profiles indexed to past evolution success record
Physical assay signal database and success indexes
select from pre-defined GO sets and profiles
Yes
No
library of GO sets and profiles
Save
Load
modify existing GO set and parameter profile
create new GO set and parameter profile
select set GOs from libraries of elemental GOs
define parameters of selected GOs
weight/ frequency
randomness
Elitism criteria
library of end criteria
define GO process end criteria
define GO application process tree structure
build a tree
library of trees
Fitness criteria library
Homology specific
Structure-based
Other (define elitism)
Sequence-specific
Parent-specific
Position-specific
GO library (lib. of objects, open architecture)
Library of multiplication GOs with default settings
Library of mutation GOs with default settings
Duplication GO
Replication GO
Other multiplication GO
Multiplication GO
single character replacement GO
clustered character replacement GO
Codon triplet replacement GO
Other clustered replacement GO
Mutation GO
single character deletion mutation GO
clustered character deletion mutation GO
single character insertion mutation GO
clustered character insertion mutation GO
other mutation GO
Fragmentation GO
Library of fragmentation GOs with default settings
Crossover/Recombination GO
Library of crossover GOs with default settings
Death GO
Other GO

Figure 3

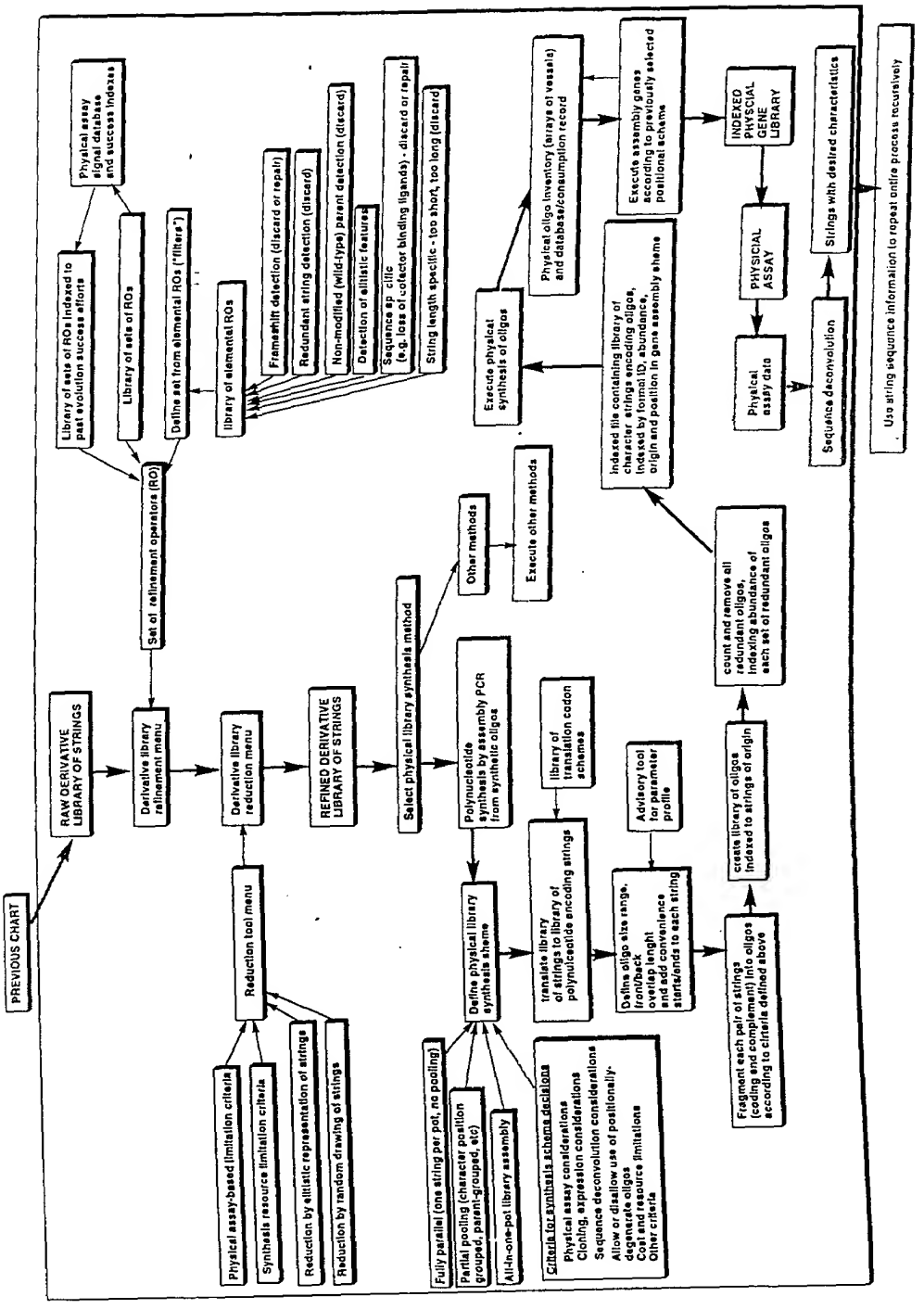

PREVIOUS CHART
RAW DERIVATIVE LIBRARY OF STRINGS
Derivative library refinement menu
Derivative library reduction menu
REFINED DERIVATIVE LIBRARY OF STRINGS
Select physical library synthesis method
Other methods
Execute other methods
Polynucleotide synthesis by assembly PCR from synthetic oligos
Define physical library synthesis sheme
Translate library of strings to library of polynucleotide encoding strings
Define oligo size range, front/back overlap lenght and add convenience starts/ends to each string
Fragment each pair of strings (coding and complement) into oligos according to criteria defined above
create library of oligos indexed to strings of origin
count and remove all redundant oligos, indexing abundance of each set of redundant oligos
Indexed file containing library of character strings encoding oligos, indexed by formal ID, abundance, origin and position in gene assembly sheme
Execute physical synthesis of oligos
Physical oligo inventory (arrays of vessels) and database/consumption record
Execute assembly genes according to previously selected positional scheme
INDEXED PHYSICAL GENE LIBRARY
PHYSICIAL ASSAY
Physical assay data
Sequence deconvolution
Strings with desired characteristics
Use string sequence information to repeat entire process recursively
Reduction tool menu
Physical assay-based limitation criteria
Synthesis resource limitation criteria
Reduction by ellistic representation of strings
Reduction by random drawing of strings
Fully parallel (one string per pot, no pooling)
Partial pooling (character position grouped, parsit-grouped, etc)
All-in-one-pot library assembly
Criteria for synthesis scheme decisions
Physical assay considerations
Cloning, expression considerations
Sequence deconvolution considerations
Allow or disallow use of positionally-degenerate oligos
Cost and resource limitations
Other criteria
library of translation codon schemes
Advisory tool for parameter profile
Set of refinement operators (RO)
Library of sets of ROs indexed to past evolution success efforts
Library of sets of ROs
Define set from elemental ROs ("filters")
library of elemental ROs
Frameshift detection (discard or repair)
Redundant string detection (discard)
Non-modified (wild-type) parent detection (discard)
Detection of ellistic features
Sequence spc cific (e.g. loss of collector binding ligands) - discard or repair
String length specific - too short, too long (discard)
Physical assay signal database and success indexes

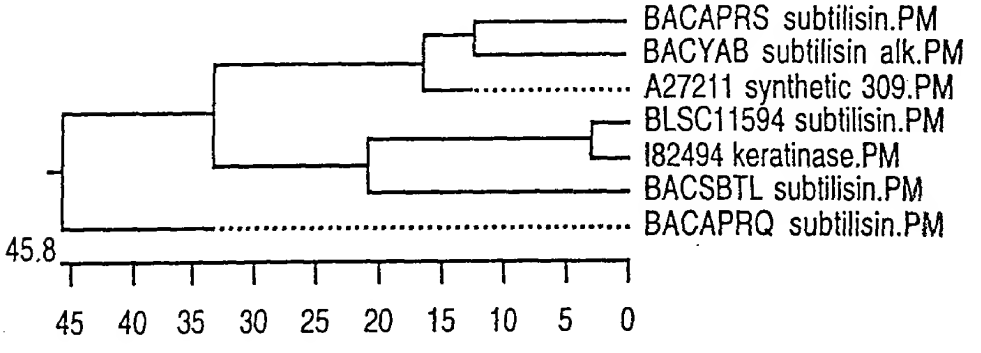
Figure 4

Figure 5

Percent Similarity

	1	2	3	4	5	6	7	
1		62.1	81.4	57.6	81.8	56.1	59.1	1
2	50.5		61.0	54.9	59.5	58.2	60.8	2
3	21.0	52.0		54.6	78.4	50.6	53.2	3
4	54.4	63.3	62.3		52.0	64.6	67.9	4
5	20.5	54.9	25.1	65.6		53.9	56.5	5
6	58.6	56.6	72.2	44.2	63.4		94.9	6
7	52.5	51.4	66.0	38.5	57.8	4.9		7
	1	2	3	4	5	6	7	

1 A27211 synthetic 309.PM
2 BACAPRQ subtilisin.PM
3 BACAPRS subtilisin.PM
4 BACSBTL subtilisin.PM
5 BACYAB subtilisin alk.PM
6 BLSC11594 subtilisin.PM
7 I82494 keratinase.PM



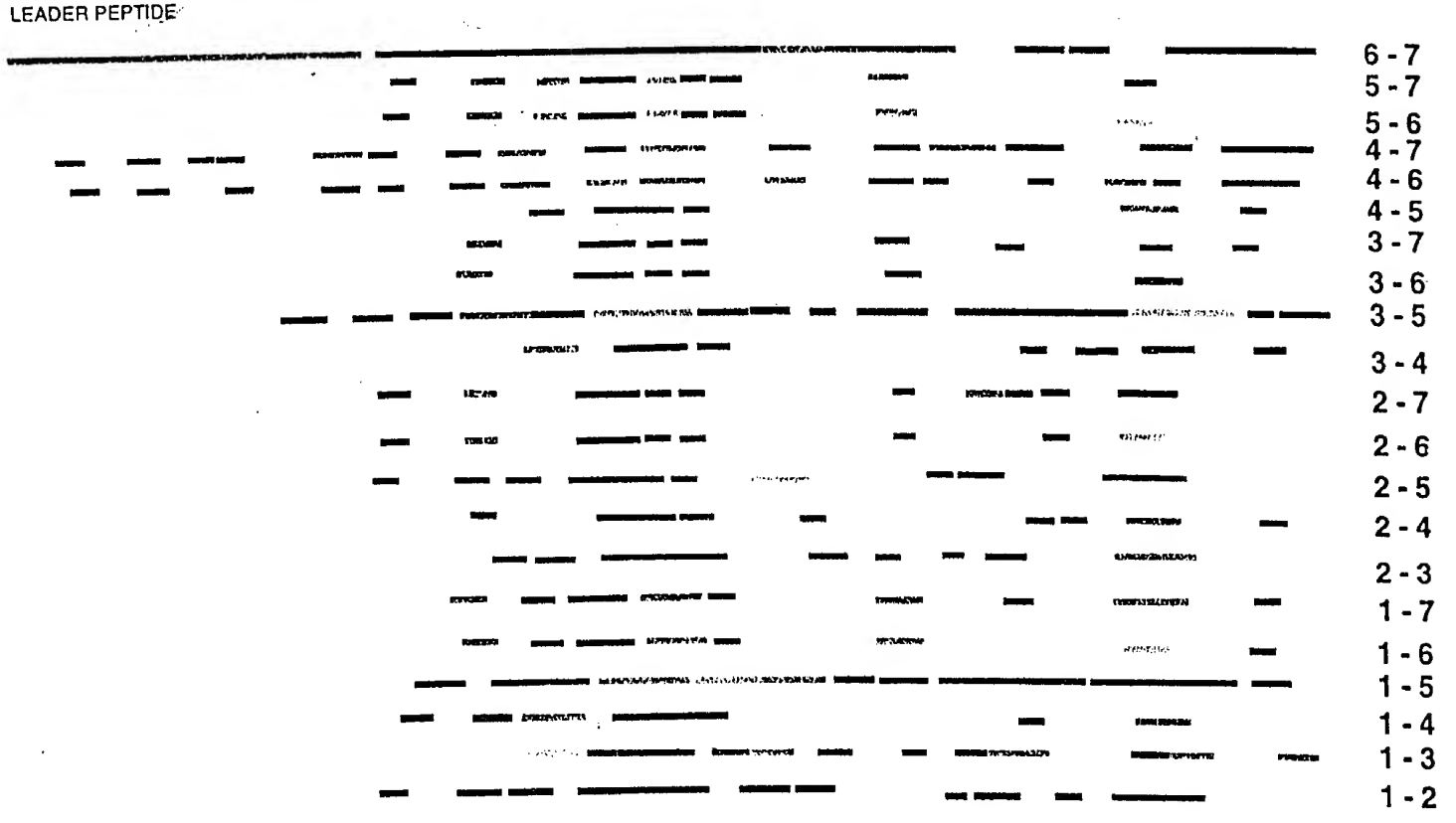

LEADER PEPTIDE
6 - 7
5 - 7
5 - 6
4 - 7
4 - 6
4 - 5
3 - 7
3 - 6
3 - 5
3 - 4
2 - 7
2 - 6
2 - 5
2 - 4
2 - 3
1 - 7
1 - 6
1 - 5
1 - 4
1 - 3
1 - 2

Figure 6

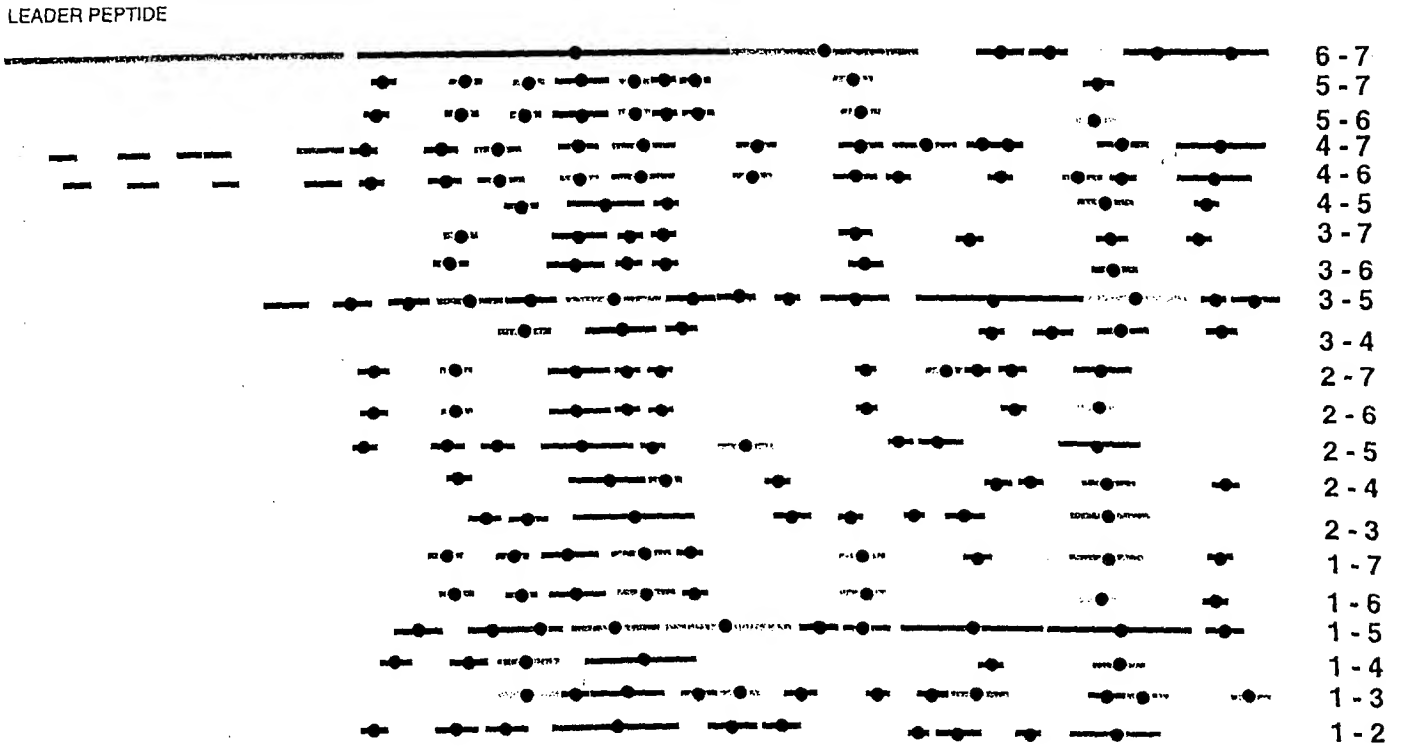

LEADER PEPTIDE
6 - 7
5 - 7
5 - 6
4 - 7
4 - 6
4 - 5
3 - 7
3 - 6
3 - 5
3 - 4
2 - 7
2 - 6
2 - 5
2 - 4
2 - 3
1 - 7
1 - 6
1 - 5
1 - 4
1 - 3
1 - 2

Figure 7

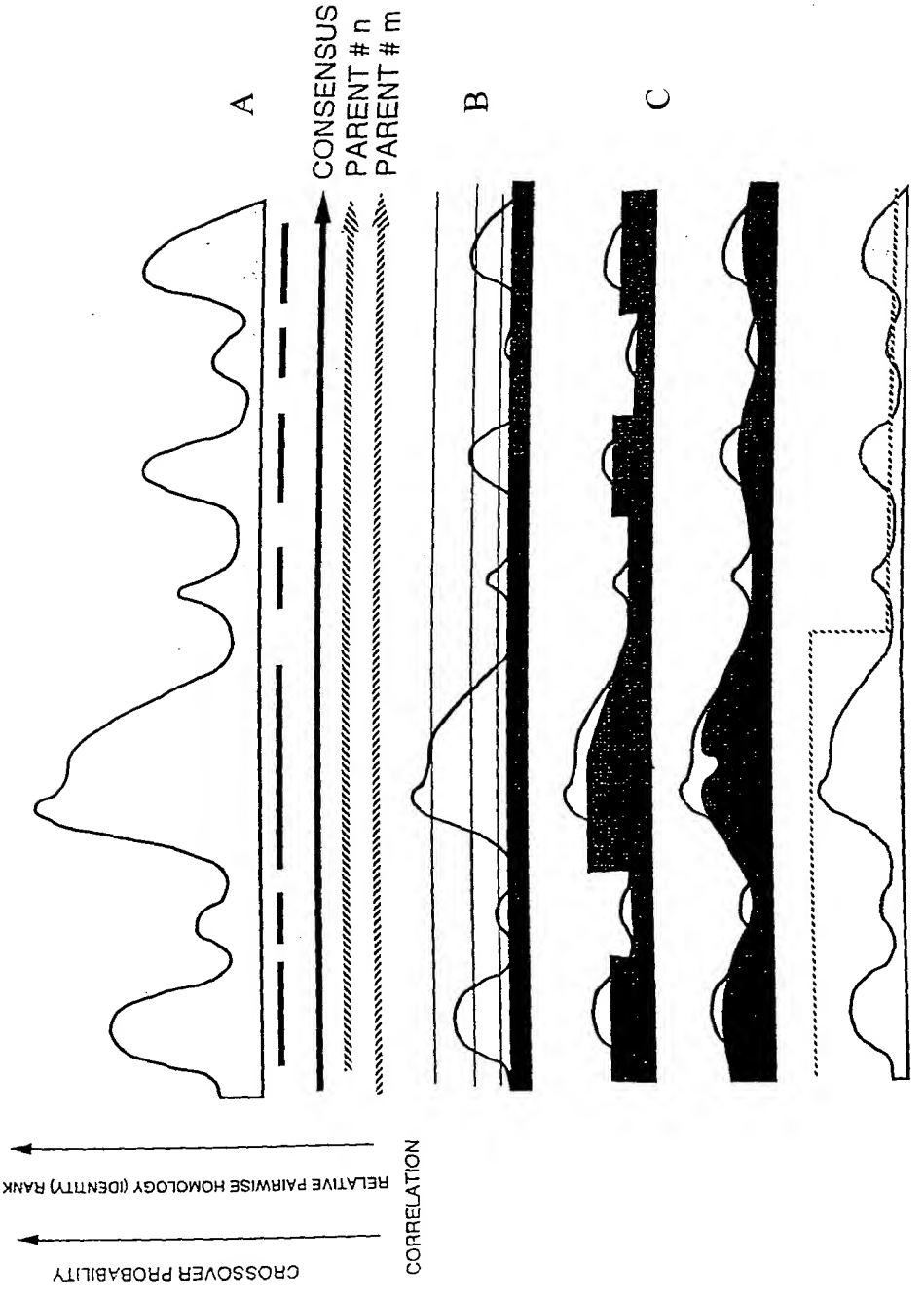

A
CONSENSUS
PARENT # n
PARENT # m
B
C
CROSSOVER PROBABILITY
RELATIVE PAIRWISE HOMOLOGY (IDENTITY) RANK
CORRELATION

Figure 8

Figure 9

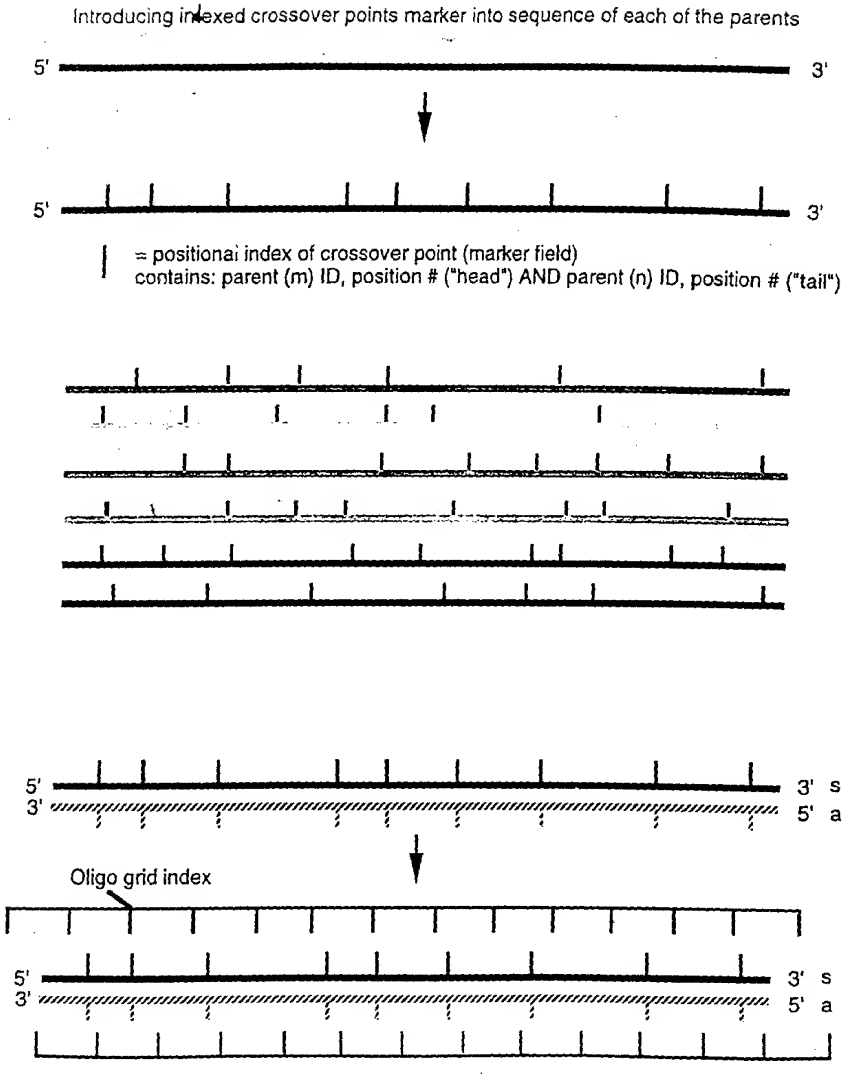

Introducing indexed crossover points marker into sequence of each of the parents
5'
3'
5'
3'
= positional index of crossover point (marker field)
contains: parent (m) ID, position # ("head") AND parent (n) ID, position # ("tail")
5'
3'
3' s
5' a
Oligo grid index
5'
3'
3' s
5' a

Figure 10

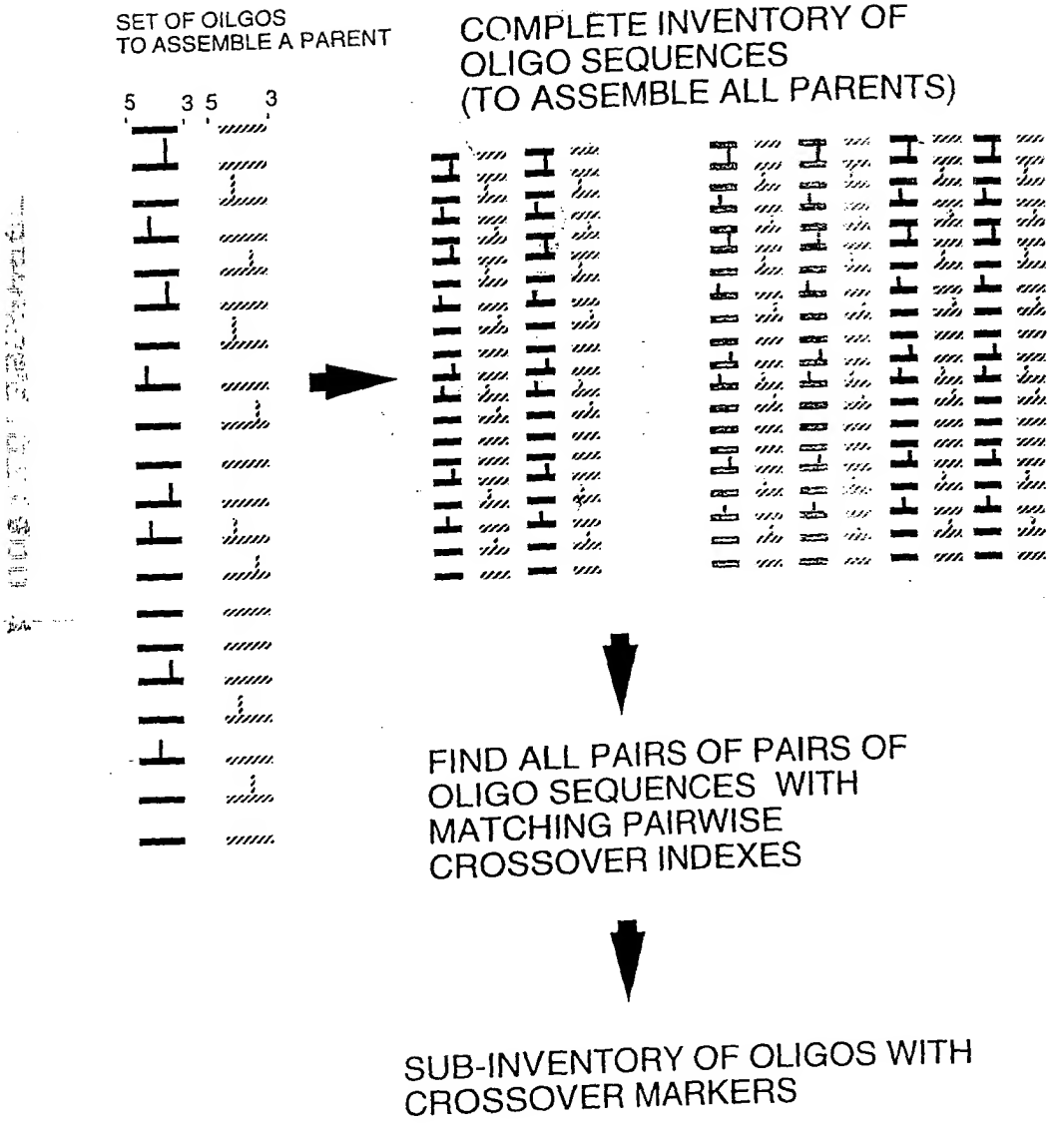

SET OF OILGOS
TO ASSEMBLE A PARENT
5 3 5 3
COMPLETE INVENTORY OF
OLIGO SEQUENCES
(TO ASSEMBLE ALL PARENTS)
FIND ALL PAIRS OF PAIRS OF
OLIGO SEQUENCES WITH
MATCHING PAIRWISE
CROSSOVER INDEXES
SUB-INVENTORY OF OLIGOS WITH
CROSSOVER MARKERS

Figure 11

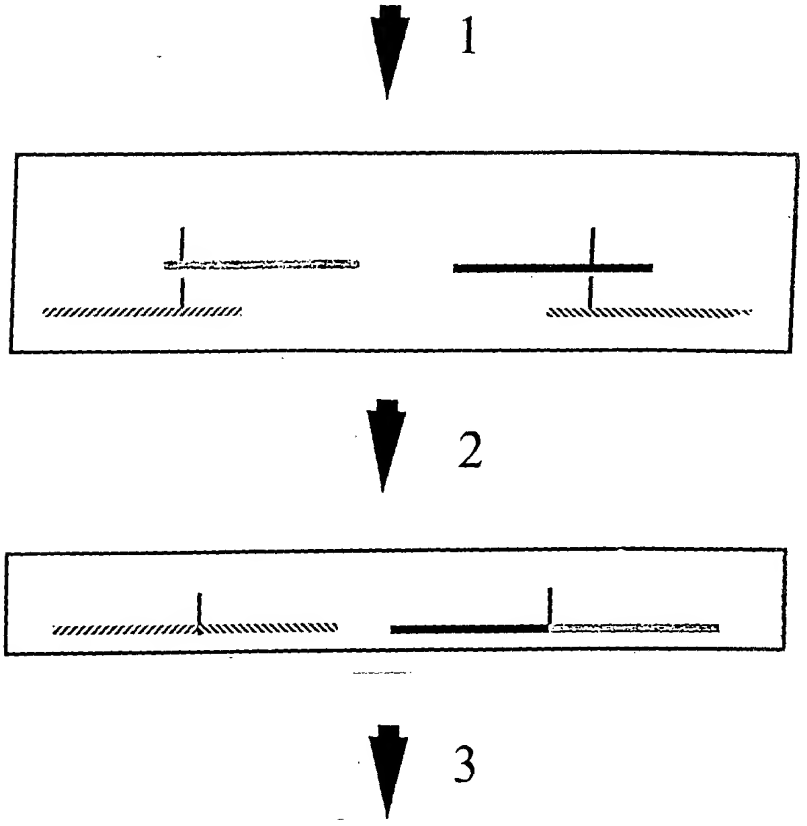


Figure 12

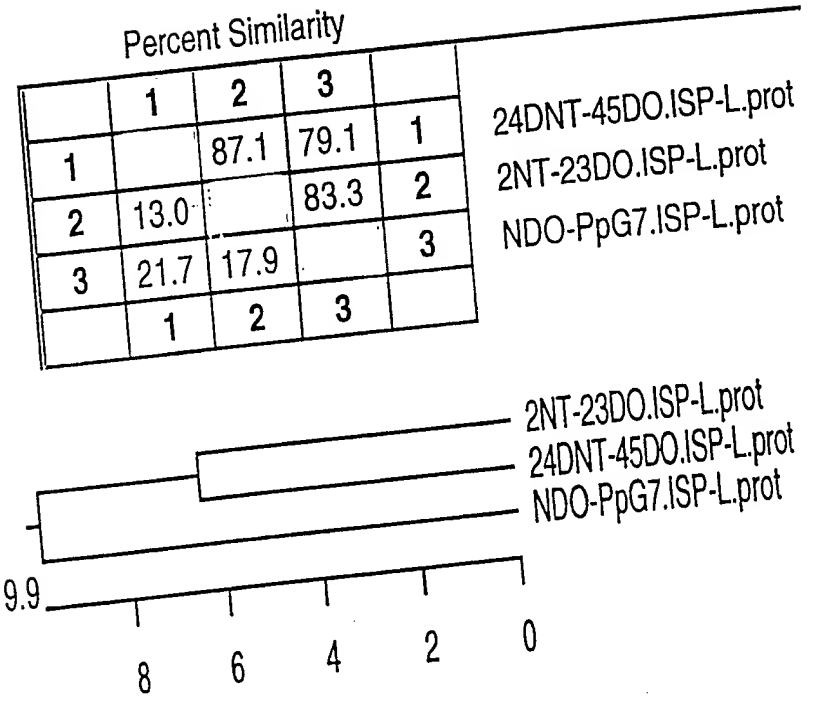

Percent Similarity
1 2 3
1 87.1 79.1 1
2 13.0 83.3 2
3 21.7 17.9 3
1 2 3
24DNT-45DO.ISP-L.prot
2NT-23DO.ISP-L.prot
NDO-PpG7.ISP-L.prot
2NT-23DO.ISP-L.prot
24DNT-45DO.ISP-L.prot
NDO-PpG7.ISP-L.prot
9.9
8
6
4
2
0

Figure 13

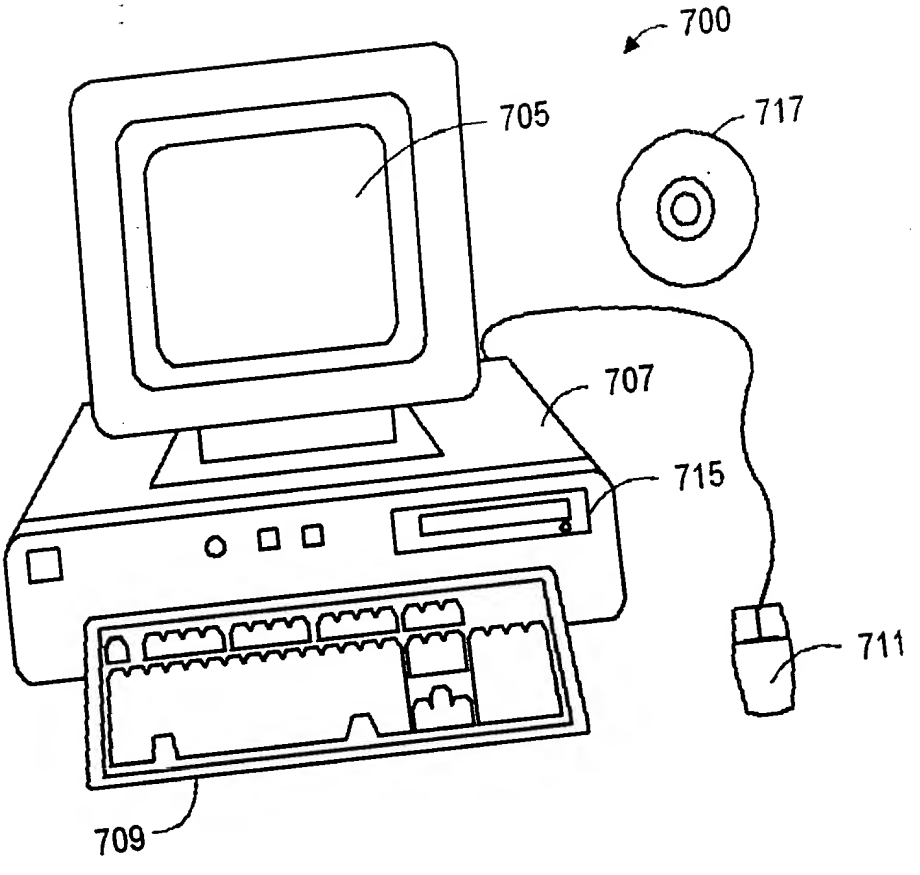
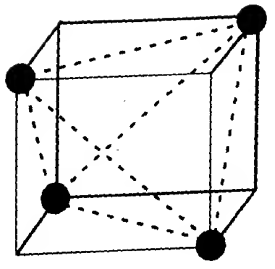


Figure 14



Geometric representation
Equal distance between
each nucleotide

A	-1 -1 +1
C	+1 -1 -1
G	-1 +1 -1
T	+1 +1 +1

Numeric representation

Figure 15